

# Parameter Estimation from an Optimal Projection in a Local Environment

A. Bijaoui, A. Recio-Blanco and P. de Laverny

*Observatoire de la Côte d'Azur BP4229  
06304 Nice Cedex 4, France*

**Abstract.** The parameter fit from a model grid is limited by our capability to reduce the number of models, taking into account the number of parameters and the non linear variation of the models with the parameters. The Local MultiLinear Regression (LMLR) algorithms allow one to fit linearly the data in a local environment. The MATISSE algorithm, developed in the context of the estimation of stellar parameters from the Gaia RVS spectra, is connected to this class of estimators. A two-steps procedure was introduced. A raw parameter estimation is first done in order to localize the parameter environment. The parameters are then estimated by projection on specific vectors computed for an optimal estimation. The MATISSE method is compared to the estimation using the objective analysis. In this framework, the kernel choice plays an important role. The environment needed for the parameter estimation can result from it. The determination of a first parameter set can be also avoided for this analysis. These procedures based on a local projection can be fruitfully applied to non linear parameter estimation if the number of data sets to be fitted is greater than the number of models.

**Keywords:** Classification, Linear Regression, Kernel method, Parameter Estimation

**PACS:** 95.75-z, 95.75.Fg, 97.10.Ri

## MODELS AND PARAMETERS.

The comparison between observations and models obtained from simulations is nowadays a common activity in astrophysics. Each model is defined by a set of physical parameters  $\Theta$ . It carries out a set of data which can be reduced to equivalent observational measurements  $S$ , taking into account the properties of the used instrumentation. Detector noise and instrumental uncontrolled variations lead to associate to observational data  $O$  a conditional probability density function (PDF)  $p(O/S)$ . Currently, this PDF is admitted to be Gaussian.

The application of the Maximum Likelihood (ML) principle [9] leads to get the parameter values  $\hat{\Theta}$  which maximize  $p(O/S)$ . Taking into account the hypothesis of a Gaussian noise, this is equivalent to determine the parameters which minimize a quadratic distance between the observations and their simulations. A large literature exists on the least mean squares (LMS) method [18]. In case of an analytical model, normal equations are derived and the problem is reduced to a classical numerical problem. Generally, the models carry out equivalent observational data which cannot be expressed as analytical functions of the physical parameters. Each comparison needs to compute a model, which can be very CPU time consuming. In order to reduce the cost, the models are computed on a parameter grid, its sampling being as large as possible such that the parameter extraction results from an interpolation in this space.

In the framework of the determination of the atmospheric parameters from the stellar spectra obtained with the Gaia RVS [17] we developed a new algorithm called MATISSE [16]. This algorithm is based on the determination of the parameter set by projection of the observed spectrum  $O$  to a set of vectors  $B_i$  computed for each parameter  $\theta_i$ . MATISSE belongs to a class of statistical methods which allows one to estimate parameters from an optimal projection on a local environment. The projection vectors derive from a multilinear regression [6] which takes only into account the models locally computed.

## THE DISTANCE MINIMIZATION AND THE MATISSE METHOD.

The parameter estimation problem consists into finding the minimum distance  $D(\Theta)$  between the observable  $O$ , here an observed spectrum, and a peculiar model from a theoretical grid, here a theoretical spectrum  $S$ . In order to simplify the presentation it is admitted that the noise variance is constant and that the measurements are independent. Therefore, the considered distance is:

$$D(\Theta) = \sum_{l=1,L} [O(l) - S(l, \Theta)]^2. \quad (1)$$

The resulting normal equation for a given parameter  $\theta_i$  can be derived. The related equation system is generally non linear. To solve it, one needs to start from a given parameter set  $\Theta^{(0)}$  and to linearize it around these values. The corrections can be large so that an iteration could be necessary. There is no guarantee that the values after convergence correspond to the absolute minimum distance. The raw solving way would consist into computing a set of models with a sufficient grid resolution for avoiding non convexity of the distance function in a given mesh. All the distances to the models are computed and the absolute minimum is localized. In a convex region the distance function is quite a quadratic function of the parameters and a simple linearization is sufficient. However, for models associated to many parameters which can vary on a large dynamical range, this procedure could be too heavy. Many algorithms were proposed for avoiding the computation of the whole model grid. Since the 70's, genetic algorithms were recognized as a fine way for solving optimization problems [8]. This technique was applied for stellar parameter estimation from spectrograms [1]. Simulated annealing [10] was also developed and appeared to be well adapted to carry out parameter estimation [5]. Artificial neural networks (ANN) were also used for this specific problem [2].

Let us come back to the normal equations. The parameter corrections can be written as:

$$\sum_{l=1,L} \frac{\partial S(l, \Theta^{(n)})}{\partial \theta_i} [O(l) - S(l, \Theta^{(n)})] = \sum_{i'=1,I} \delta \theta_{i'} \sum_{l=1,L} \frac{\partial S(l, \Theta^{(n)})}{\partial \theta_i} \frac{\partial S(l, \Theta^{(n)})}{\partial \theta_{i'}}. \quad (2)$$

So we can write the relation:

$$\delta \Theta = G^{-1} [T(O - S)]. \quad (3)$$

where  $O - S$  designs the residual vector (length  $L$ ),  $T$  the  $L \times I$  model derivative matrix and  $G$  the  $I \times I$  Gram-Schmidt matrix of the model derivatives. With Equation (3), the

parameter corrections are expressed as linear combinations of the model residuals. The associated coefficients depend on the partial derivatives. In our studied problem only computed models on a grid are available. Approximated derivatives can be computed but it is better to embed this approximation into a more general scheme taking into account the grid sampling. In this scheme, the MATrix Inversion for Spectral SynthEsis (MATISSE [16]) method, the algorithm determines a vector,  $B_i$ , allowing one to derive a particular parameter  $\theta_i$  by projection of an input observed spectrum on it. The  $B_i$  vector is derived from an optimal linear combination of models in a local environment, the environment depending on the model linearity versus the parameters.  $B_i$  is determined from the principle that the statistical correlation between the input and the output parameter values is maximum. With a spectra energy normalization this is equivalent to minimize the distance between the input and the output parameter values.

First of all, the data on a particular  $\theta_i$  parameter and the spectra of the grid are subtracted by their mean values. The  $B_i$  vector is then constructed from:

$$B_i(l) = \sum_{j=1,J} \alpha_{ij} S_j(l). \quad (4)$$

$\theta_i$  is estimated by the spectrum projection:

$$\hat{\theta}_i = \sum_l B_i(l) O(l). \quad (5)$$

In other words,  $\Theta$  is estimated by multilinear regressions done on the spectrum values. Equation (4) is applied to all the model spectra  $S_j(l)$ . Combining with Equation (5) we obtain :

$$\Theta_i = C \alpha_i, \quad (6)$$

where  $C = [c_{jj'}]$  is the correlation matrix of the spectra and  $\Theta_i$  the vector of the parameter  $i$  for all the spectra. For an invertible variance-covariance matrix, we get:

$$\alpha_i = C^{-1} \Theta_i, \quad (7)$$

which leads for the training set to:

$$\hat{\Theta}_i = SB = S(S\alpha_i) = CC^{-1} \Theta_i = \Theta_i. \quad (8)$$

The exact values are restored. In order to be sure that a peculiar model was not forgotten for the basis building a redundancy is necessary. That leads to a non invertible matrix. The LMS solution between the input parameters and their estimated ones leads to:

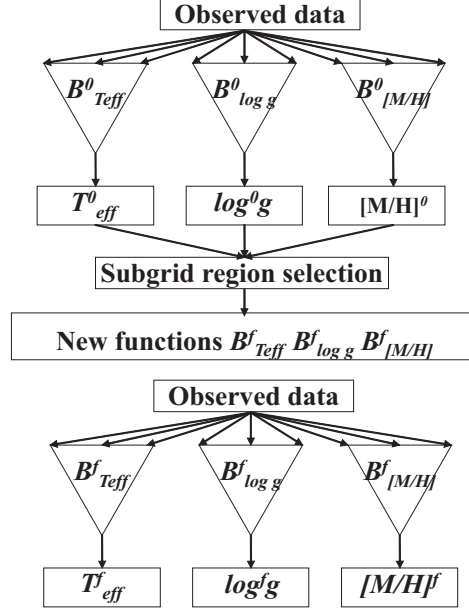
$$C^T C \alpha_i = C^T \Theta_i. \quad (9)$$

Even if  $C$  is singular, the multiplication by its joint matrix  $C^T$  brings a solution in the LMS sense. The inversion can be done, for example, using the classical Landweber iterative algorithm (Landweber, 1951).

The multilinear regression model stands only if the model is quite linear. Generally it is not the case and it is needed to process in two steps (Fig. 1):

1. A preliminary estimation of the parameters is performed through the use of  $B_{\Theta}^{(0)}$  functions computed from spectra combinations spanning the parameter range of the whole grid.

2. This first estimation allows one to focus into a subregion of the model grid, where new  $B_{\Theta}^{(f)}$  functions constructed from a subset of theoretical spectra are applied.



**FIGURE 1.** Scheme of the MATISSE algorithm for the estimation of the stellar parameters.

## THE OBJECTIVE ANALYSIS WITH KERNEL.

The *Objective Analysis* recovers a large set of interpolation methods developed in the framework of Geosciences. The function  $f(x)$  is known on  $J$  points  $(x_j, j \in (1, J))$  and it is interpolated with the rule:

$$f(x) = \frac{\sum_j w_j(x) f(x_j)}{\sum_j w_j(x)} \quad (10)$$

Many determinist or statistical methods, like the kriging [14] [12], were proposed for the determination of the coefficients  $w_j(x)$ . The weights can be directly obtained from a Kernel function  $K(\frac{x}{a})$  [11],  $a$  being a scaling parameter. The Nadaraya-Watson weighted average [7] consists in the following rule:

$$\hat{f}(x) = \frac{\sum_j K(\frac{|x-x_j|}{a}) f(x_j)}{\sum_j K(\frac{|x-x_j|}{a})}. \quad (11)$$

In order to obtain a true interpolation for which the values are exactly recovered at the mesh points, the set  $\bar{f}(x_j)$  such that:

$$f(x_j) = \frac{\sum_{j'} K\left(\frac{|x_j - x_{j'}|}{a}\right) \bar{f}(x_{j'})}{\sum_{j'} K\left(\frac{|x_j - x_{j'}|}{a}\right)} \quad (12)$$

is determined. Equation (12) is solved iteratively. The interpolation rule on the parameters  $\theta_i$  is thus:

$$\hat{\theta}_i(O) = \frac{\sum_j K\left(\frac{|O - S_j|}{a}\right) \bar{\theta}_{ij}}{\sum_j K\left(\frac{|O - S_j|}{a}\right)}, \quad (13)$$

where the  $\bar{\theta}_{ij}$  are previously determined by inversion.

Let us consider the Epanechnikov kernel [3]:

$$K(x) = \frac{3}{4}(1 - x^2) \quad \text{if} \quad |x| \leq 1 \quad \text{otherwise} \quad K(x) = 0. \quad (14)$$

Taking into that the spectra normalization the weights can be simplified as:

$$w_{jj'} = H(c_{jj'} - c), \quad (15)$$

where  $c$  is a correlation threshold and  $H(x)$  the Heaviside function ( $H(x) = 1$  if  $x \geq 0$  and null otherwise). The weights for a spectrum  $O$  are thus:

$$W_j = H(c_j - c) = H\left(\sum_l O(l) S_j(l) - c\right). \quad (16)$$

The parameter  $\theta_i$  is estimated by the relation:

$$\hat{\theta}_i = \frac{\sum_j H(\sum_l O(l) S_j(l) - c) \bar{\theta}_{ij}}{\sum_j H(\sum_l O(l) S_j(l) - c)}. \quad (17)$$

Defining  $e(O)$  as the set of spectra  $S_j$  such that  $c_j > c$ , relation (17) can be written as:

$$\hat{\theta}_i = \frac{\sum_l O(l) [\sum_{j \in e(O)} S_j(l) \bar{\theta}_{ij}] - c \sum_{j \in e(O)} \bar{\theta}_{ij}}{W}, \quad (18)$$

where  $W = \sum_{j \in e(O)} W_j$ . Thus it results that:

$$\hat{\theta}_i = \sum_l O(l) B_i^e(l) - \theta_i^e \quad (19)$$

where  $B_i^e(l) = \frac{1}{W} \sum_{j \in e(O)} S_j(l) \bar{\theta}_{ij}$  and  $\theta_i^e = \frac{c \sum_{j \in e(O)} \bar{\theta}_{ij}}{W}$ . The parameters  $\bar{\theta}_{ij}$  are determined by the inversion of the equation:

$$\theta_{ij} = \frac{\sum_{j'} H(c_{jj'} - c) \bar{\theta}_{ij}}{\sum_j H(c_{jj'} - c)}. \quad (20)$$

**TABLE 1.** Experimental results on a reduced grid.

Method	$T$ -max	$\sigma(T)$	$\log g$ -max	$\sigma(\log g)$	$\mu$ -max	$\sigma(\mu)$
Kernel at 0.995	122	13	0.42	0.06	0.17	0.03
MATISSE B0	1402	451	2.95	0.69	1.32	0.26
MATISSE B0+Bf	1709	104	6.	0.19	1.34	0.08
B0+Bf 99% stars	0.	0.	0.	0.	0.	0.

This equation, with a null threshold, is equivalent to the one associated to the MATISSE method, giving the  $\alpha_i$  value (Equation 6). In the MATISSE case, the environment is defined from the grid sampling. Here the environment results from the correlation sampling for the objective analysis. The estimation is done by a correlation product, with a function which varies according to the correlations. It can be noted that the  $B_i^e$  varies for a spectrum to another, but the contributions associated to each spectrum are computed in one step.

## EXPERIMENTATIONS ON A SPECTRA SET.

A set of 1386 spectra with 971 Gaia RVS spectral elements were computed. In this grid, 3 physical parameters vary, the temperature  $T$ , the surface gravity  $\log g$  and the metallicity  $\mu$ . Experiments were done with different kernels. In the case of the Epanechnikov kernel, different correlation thresholds were examined. For the studied grid, the environment, defined by the number of remaining spectra for each one, seems to be sufficient at a 0.995 threshold. In Table 1 a comparison is presented with different experiments:

**Kernel** The parameters are restored using the Nadaraya-Watson weighting average, with the Epanechnikov kernel. The correlation threshold is equal to 0.995. The environment varies from 2 to 68 spectra with this value. The minimum number is obtained at points on the grid bound, for which the environment is insufficient to get an accurate value.

**MATISSE B0** The parameters are determined by one projection. The maximal errors are greater than the grid steps. This operation allows one only to localize the environment for obtaining the correct values.

**MATISSE B0+Bf** After localization, a projection is done for obtaining correct values. Few iterations may be needed in order to get the environment corresponding to the parameters.

**MATISSE with rejection** The previous algorithm works perfectly on 99% of the grid spectra. The errors are related to the grid boundary. At the first step the estimated parameters may be too far from real ones. The  $B_f$  corrections do not allow us to converge correctly. A simple test based on the correlation between the restored spectrum and the observed one allows us to reject the bad estimations.

These experiments show that the kernel method is well designed for a first estimation of the parameter values on a reduced grid. It is necessary for that to compute theoretically

all the correlations. Specific algorithms have to be developed for reducing this number taking into account the correlation threshold.

## CONCLUSION.

The MATISSE method allows one to improve model fitting in the case of:

- A capability to describe observations by a suitable model depending of few parameters;
- The models can not be analytically manipulated, avoiding to use their derivatives for the estimations;
- The need to fit a huge number of observational data.

The main part of the computation is related to the basis estimation, the parameter estimation itself being very fast. The use of the Epanechnikov's kernel method allows one to get in one inversion all the coefficients. The number of spectra having a correlation above the threshold according to a given grid point indicates if locally the sampling is in agreement with the searched accuracy.

Taking into account their capabilities to process many sets of astronomical observations, with a high statistical consistence level, the estimation methods based on local projections like MATISSE open a new window on astrophysical modelization for large astrophysical surveys.

## REFERENCES

1. C. Allende Prieto, *MNRAS* **339**,1111–1116 (2003).
2. C.A.L. Bailer-Jones, , M. Irwin, , G. Gilmore and T. Von Hippel, *MNRAS* **292**, 157–166 (1997).
3. V.A. Epanechnikov, *Theory Probab. Appl.* **14**, 153–158 (1969).
4. L. Gandin, *Objective Analysis of Meteorological Fields*, Leningrad, 1963.
5. O.R. Gonzalez, , C. Küper, K. Jung, C. Naval and E. Mendoza, E., *Bioinformatics* **23**, 480–486 (2007).
6. T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*,(2001), Springer, Francfort, p.41.
7. T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*,(2001), Springer, Francfort, p.35.
8. J.H. Holland, *Adaptation in Natural and Artificial Systems*, University of Michigan Press, Ann Arbor, 1975.
9. M.G. Kendall and A. Stuart, *The Advanced Theory of Statistics II*, Griffin, London, 1973, p.8.
10. S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, *Science* **200**, 671–680 (1983).
11. S.E.Koch, M.DesJardins and P.J.Kocin, *Jour. of Climate and Applied Meteorology* **22**, 1487–1503 (1983).
12. D.G. Krige, “A review of the development of geostatistics in South Africa,” in *Advanced geostatistics in the mining industry*, Reidel: Dordrecht, Holland, 1976.
13. L. Landweber, *Am. J. Math.* **73**, 615–624 (1951).
14. G. Matheron, *Les variables régionalisées et leur estimation*, Masson, Paris, 1975.
15. A. Recio-Blanco, P. de Laverny and B. Plez, *ESA Technical note*, RVS-ARB-001 (2005).
16. A. Recio-Blanco, A. Bijaoui and P. de Laverny, *MNRAS* **370**, 141–150 (2006).
17. M.I. Wilkinson and 40 authors, *MNRAS* **359**, 1306–1335 (2005).
18. J. Wolberg, *Data Analysis Using the Method of Least Squares: Extracting the Most Information from Experiments*, Springer, 2005.